ORIGINAL

BELLSOUTH

BellSouth
Suite 900
1133-21st Street, N.W.
Washington, D.C. 20036-3351

kathleen.levitz@bellsouth.com

Kathleen B. Levitz
Vice President-Federal Regulatory

202 463-4113
Fax 202 463-4198

EX PARTE OR LATE FILED

December 7, 1999

RECEIVED

DEC - 7 1999

FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

EX PARTE

Ms. Magalie Roman Salas
Secretary
Federal Communications Commission
The Portals
445 12th St. SW
Washington, D.C. 20554

Re: CC Docket No. 98-56 and CC Docket No. 98-121

Dear Ms. Salas:

This is to inform you that on December 6, 1999 Venetta Bridges and I, representing BellSouth, and Dr. Fritz Scheuren and Dr. Edward Mulrow of Ernst & Young met with Daniel Shiman, John Stanley, Andre Rausch, Raj Kannan, and Alex Belinfante of the Common Carrier Bureau. During the meeting we discussed the continuing efforts of the statisticians at Ernst and Young to develop a method of statistical analysis that the Louisiana Public Service Commission in LPSC Docket No. U22252 – Subdocket C could approve for use in determining whether BellSouth is meeting its statutory obligation to provide CLECs with nondiscriminatory access to UNEs and services. The attached documents formed the basis for that discussion.

Because the Commission has been considering issues related to performance measurements and standards in both proceedings identified above, we are filing

notice of this ex parte meeting in both dockets, as required by Section 1.1206(b)(2) of the Commission's rules. Please associate this notice with the record of both dockets.

Sincerely,

Kathleen B. Levitz

Attachments

cc: Daniel Shiman (w/o attachments)
John Stanley (w/o attachments)
Andre Rausch (w/o attachments)
Raj Kannan (w/o attachments)
Alex Belinfante (w/o attachments)

LPSC October Workshop

Statistician's Report

Original Consensus/Open Issues

and

Resolutions

October 25, 1999

Original Consensus/Open Issues Matrix - April 1999

Issue No.	Issue	Position
1	Comparing like-to-like	**Agreement:** In order to assure that like-to-like comparisons are made, the performance measure data must be disaggregated to a very deep level. This includes wire center and time of month, as well as SQM disaggregation levels defined by the Louisiana Public Service Commission.[1]
2	Performance measure test statistic	**Agreement:** Each performance measure of interest should be summarized by one overall test statistic giving the decision maker a rule that determines whether a statistically significant difference exists.
3	Methodology for obtaining the test statistic	**Dr. Mallows/LCUG**: In each cell, construct an indicator that is sensitive to absence of parity.. Make appropriate allowance for what would be the effect of random variation, assuming parity holds. The aggregate statistic should not allow consistent violations in any cell to go undetected. **BellSouth**: The overall service process is what defines parity. Testing measures at an aggregate level is sufficient to determine favoritism. Random failures at deeply disaggregated levels may exists but should not be overemphasized. SQM level disaggregation reports will be available to explore the data.
4	Type I and Type II errors	**Agreement:** The probability of a Type I error, concluding BellSouth favoritism exists when it does not, should be balanced with the probability of a type II error, concluding there is no BellSouth favoritism when there is. The balance of these two probabilities depends on 1. The effective number of BellSouth observations 2. The effective number of CLEC observations 3. The size of a specific alternative hypothesis, e.g., the CLEC mean value is larger than the BellSouth mean value by ten percent of a BellSouth standard deviation Using this information, a critical value for the test, or decision rule, is determined. This rule may be different for each performance measure in interest, and may also change over the months. However, a system can be devised to make this all transparent to the commission.

[1] Louisiana Public Service Commission Docket No. U-22252-Subdocket C, In Re: BellSouth Telecommunications Inc., Service Quality Performance Measurements, April 19, 1998 Order. Except that for provisioning measures order type was also included since there is a noticeable difference in their distributions.

Original Consensus/Open Issues Matrix - April 1999

Issue No.	Issue	Position
4a	Type I and Type II errors	**Dr. Mallows/LCUG**: We do not agree that the following BellSouth alternative is either feasible (since it requires the parties to agree on what constitutes a material difference), or fair (since it uses a test procedure at a level (2 1/2%) that is biased in favor of BellSouth for all sample sizes below 1000). **BellSouth**: If the balancing procedure described in Issue Number 4 is determined to be unworkable, then a feasible alternative is to define the size of a difference between mean values which has no business impact (a rule of materiality). Any actual difference less than this will be considered insignificant. Differences greater than the materiality standard would be judged to be significant based on a statistical testing procedure. This should be a five percent (5%) significance level, two-sided test (a two and one half percent (2.5%) significance level, one-sided test).
5	Statistical paradigm	**Agreement:** The system must be developed so that it can be put into production (black box). Two statistical paradigms are possible for examining the performance measure data. In the exploratory paradigm, data are examined and methodology is developed that is consistent with what is found. In a production paradigm a methodology is decided upon before data exploration. While the exploratory paradigm provides protection against using erroneous data it requires a great deal of lead time and is unsuitable for timely monthly performance measure testing. A production paradigm will not only promptly produce overall test results but will also provide documentation that can be used to explore the data after the test results are released.
6	Trimming	**Agreement:** Trimming is needed but finding a robust rule that can be used in a production setting is difficult. Trimming of extreme observations from BellSouth and CLEC distributions is needed in order to ensure that a fair comparison is made between performance measures. However, trimmed observations should not simply be discarded. They need to be examined and possibly used in the final decision making process. Under a production paradigm this is very hard to do. Additionally, each performance measure may need to use a different trimming rule.
7	Independence of performance measure tests	**Agreement:** Correlation between the performance measures must be accounted for in aggregation over performance measures.

Statistical Techniques For The Analysis
And Comparison Of Performance Measurement Data
September 1999

Comparing Like-to-Like

Position	Statistical Resolution
In order to assure that like-to-like comparisons are made, the performance measure data must be disaggregated to a very deep level. This includes wire center and time of month, as well as SQM disaggregation levels defined by the Louisiana Public Service Commission.[1]	• Identify variables that may affect the measure • Use a test statistic that is robust with respect ot unnecessary disagregation • Appendix B – Trunk Blocking provides an example

Statistical Techniques For The Analysis
And Comparison Of Performance Measurement Data
September 1999

Performance Measure Test Statistic Methodology

Position	Statistical Resolution
Each performance measure of interest should be summarized by one overall test statistic giving the decision maker a rule that determines whether a statistically significant difference exists. In each cell, construct an indicator that is sensitive to absence of parity. This indicator should be standardized to allow for the effect of random variation, assuming parity holds. Aggregate these standardized indicators in a way that does not allow consistent violations in any cell or group of cells to go undetected. As far as possible, cancellation should not be allowed to occur. There is complete agreement on rates and proportions. Averages pose an operational problem. Feasible methods for small sample sizes are under development.	• Test statistic calculated in each like-to-like cells • Truncate Z statistic is used to combine cells tests into one overall test statistic • Cancellation of results between cells results is limited • Details are contained in Appendix A

Statistical Techniques For The Analysis
And Comparison Of Performance Measurement Data
September 1999

Type I and Type II errors

Position	Statistical Resolution
The probability of a Type I error, concluding BellSouth favoritism exists when it does not, should be balanced with the probability of a type II error, concluding there is no BellSouth favoritism when there is. The balance of these two probabilities depends on 1. The effective number of BellSouth observations 2. The effective number of CLEC observations 3. The size of a specific alternative hypothesis, e.g., the CLEC mean value is larger than the BellSouth mean value by ten percent of a BellSouth standard deviation Using this information, a critical value for the test, or decision rule, is determined. The agreed upon balancing formula for a single test is attached. The balancing formula for an aggregated test is still under development.	• Alternative Hypotheses within each cell should be considered • Subject matter experts should work with statisticians to determine a set of alternative hypotheses • Tier I and Tier II testing may require different alternative hypotheses • Mathematical formula exists that balances the truncated Z statistic • Appendix C contains the details behind balancing

Statistical Techniques For The Analysis
And Comparison Of Performance Measurement Data
September 1999

Statistical paradigm

Position	Statistical Resolution
The system must be developed so that it can be put into production (black box). Two statistical paradigms are possible for examining the performance measure data. In the exploratory paradigm, data are examined and methodology is developed that is consistent with what is found. In a production paradigm a methodology is decided upon before data exploration. While the exploratory paradigm provides protection against using erroneous data it requires a great deal of lead time and is unsuitable for timely monthly performance measure testing. A production paradigm will not only promptly produce overall test results but will also provide documentation that can be used to explore the data after the test results are released.	• New test statistic developed for testing means that adjusts for skewness. This allows test to be used on fairly small samples, and avoids permutation tests. (Discussed in Appendix A) • When permutation testing is necessary, a fast and efficient algorithm has been developed. • Final reporting requirements should be determined by the LPSC • Appendix D contains example reports and suggests ways to report facts in a systematic way so that observers can judge if the test results are reasonable.

Statistical Techniques For The Analysis
And Comparison Of Performance Measurement Data
September 1999

Trimming

Position	Statistical Resolution
Trimming is needed but finding a robust rule that can be used in a production setting is difficult. Trimming of extreme observations from BellSouth and CLEC distributions is needed in order to ensure that a fair comparison is made between performance measures. However, trimmed observations should not simply be discarded. They need to be examined and possibly used in the final decision making process. Each performance measure may need to use a different trimming rule.	• Trimming is only necessary for mean measures • Trim BST data at the largest observed CLEC value • Report information on trimming point, BST values trimmed, and large CLEC values • Appendix E provides details

Statistical Techniques For The Analysis
And Comparison Of Performance Measurement Data
September 1999

Independence of Performance Measure Tests

Position	Statistical Resolution
Correlation between the performance measures must be accounted for in aggregation over performance measures.	This issue was not addressing in the Statistical Report

Issues to Resolve

- Alternative Hypotheses
- Benchmarks
- Test Reporting Structure
- Trunk Blocking

Suppose the alternative hypothesis is given by

H_a: $\mu_{2j} = \mu_{1j} + \delta_j \cdot \sigma_{1j}$, $\sigma_{2j}^2 = \lambda_j \cdot \sigma_{1j}^2$ $\delta_j > 0$, $\lambda_j \geq 1$ and $j = 1,\ldots,L$, where L is the total number of occupied cells.

To simplify matters let $\lambda_j = 1$, and $\delta_j = \delta$

Consider the balancing critical value

$$c_B = \frac{-\delta}{2\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where n_1 and n_2 are the aggregate ILEC and CLEC volumes

For April 99 there are 43 CLECs with maintenance troubles, and 39 CLECs with provisioning orders. This means that there will be 82 critical values for individual CLEC vs BST tests, and two more for aggregate CLEC vs BST tests. The following is a summary of the critical values when $\delta = 1$. Summaries for other deltas are scalar multiples of this one.

Table 1: Critical Values for a Delta Value of 1

Statistic	All Tests	Maintenance	Provisioning Only
Minimum	-73.70617	-32.16080	-73.70617
1st	-7.31980	-5.84804	-8.21805
Median	-3.74895	-2.73772	-6.23456
3rd Quatile	-1.58095	-1.41390	-3.21234
Maximum	-0.35355	-0.35355	-0.48507
Total N	84	44	40

Box Plots

Figure 1: Box Plot of Critical Value for All Tests ($\delta = 1$)

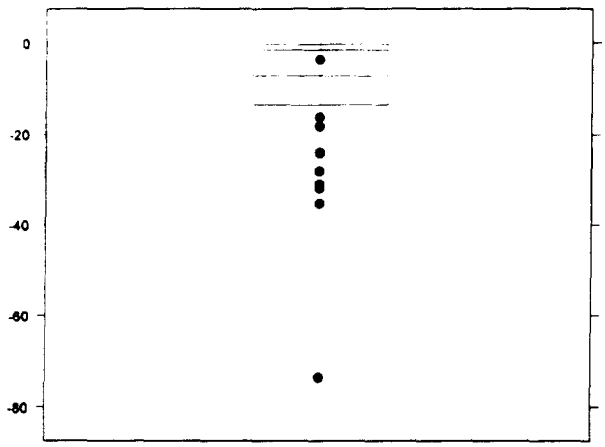
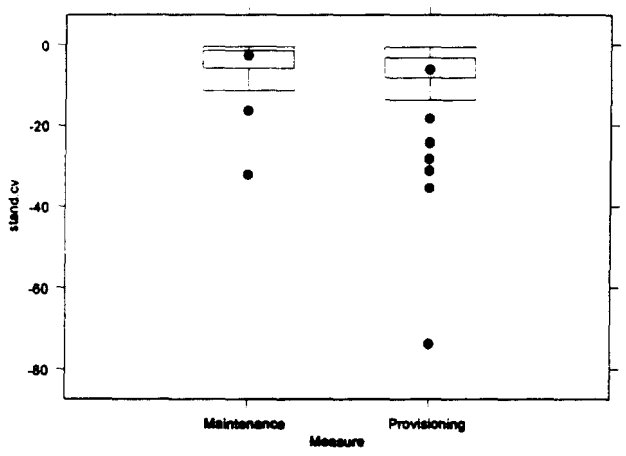
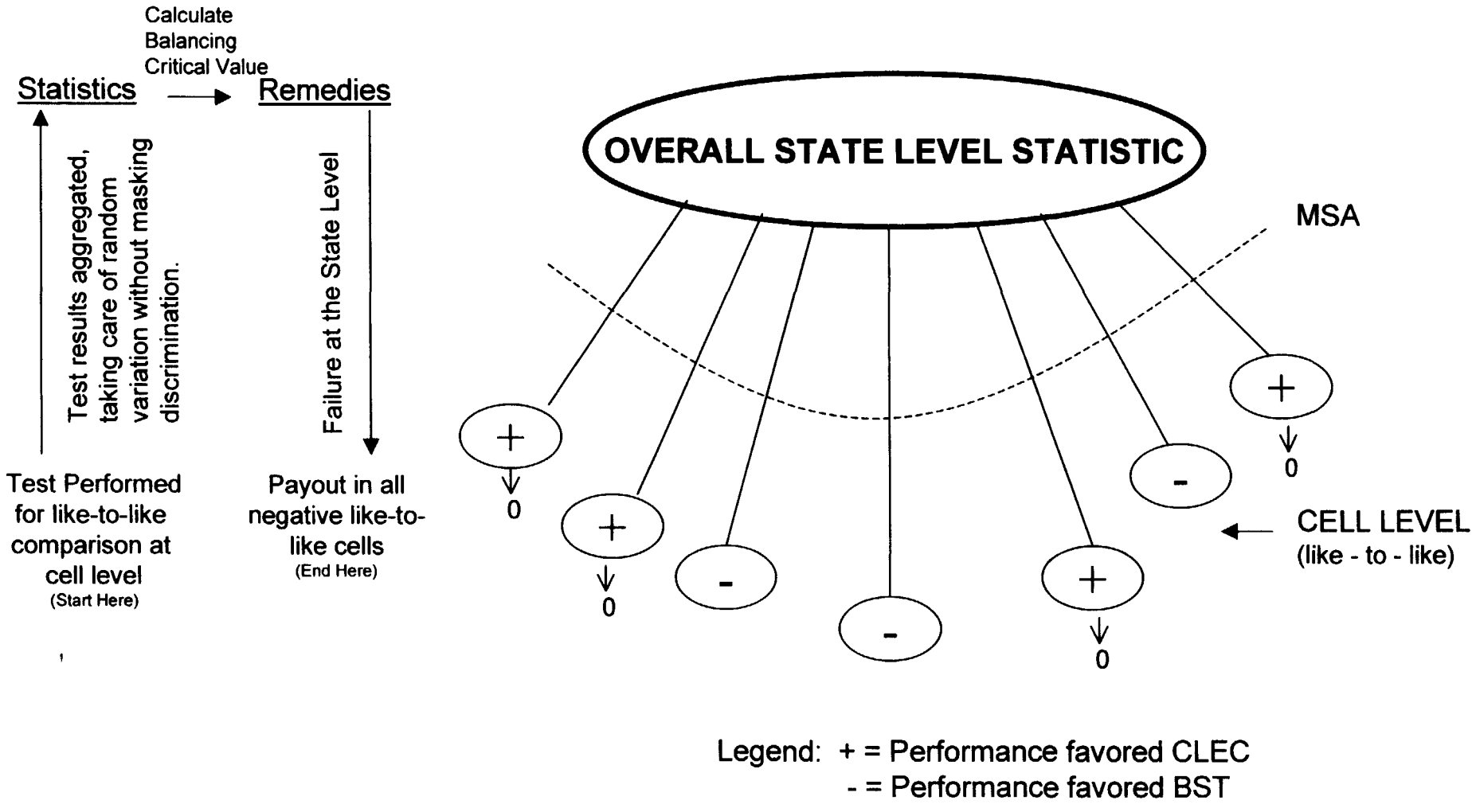


Figure 2: Box Plot of Critical Value by Type ($\delta = 1$)



Thus, the median balancing critical value is -3.75 for $\delta = 1$, but it's -0.375 for $\delta = 0.1$.

Remedy Payout Diagram



Legend: + = Performance favored CLEC
- = Performance favored BST